DOCUMENT RESUME

ED 443 827 TM 031 309

AUTHOR	Pomplun, Mark; Frey, Sharon; Becker, Douglas; Hughes, Kay
TITLE	The Validity of a Computerized Measure of Reading Rate.
PUB DATE	2000-04-00
NOTE	20p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*College Students; *Computer Assisted Testing; *High School Students; High Schools; Higher Education; Reading Comprehension; *Reading Rate; *Scores; Test Format; *Validity
IDENTIFIERS	*Nelson Denny Reading Tests

ABSTRACT

This paper describes the investigation of a computerized measure of reading rate as measured by the new Nelson-Denny Reading Test CD-ROM (1993). This study addressed three aspects of validity: (1) score comparability between reading rate measured by the computer version and the paper-and-pencil version; (2) concurrent validity with reading comprehension; (3) and predictive validity with academic achievement. Concurrent validity was studied through a score equivalence special study with 182 high school and college students taking both the computerized and paper-and-pencil versions. Mean score differences were not statistically significant, and correlations were significant at the .05 level. Construct validity was displayed through a curvilinear relationship with reading comprehension in addition to adding a significant increase to predicted variance for reading comprehension after reading vocabulary was first used as a predictor. Predictive validity was demonstrated by significant additions to the percent of variance explained in the prediction of high school and English course grades after the reading total score was first used as a predictor. The results of this study support the use of this computerized measure of reading rate and suggest several areas of possible usefulness in the understanding of reading comprehension and the prediction of academic achievement. (Contains 2 figures, 3 tables, and 15 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 443 827

The Validity of a Computerized Measure of Reading Rate

Mark Pomplun, Sharon Frey, Douglas Becker and Kay Hughes

Riverside Publishing

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

M. Pomplun

Paper presented at April 2000 convention of American Educational Research Association in New Orleans.

TM031309

BEST COPY AVAILABLE

Abstract

This paper describes the investigation of a computerized measure of reading rate as measured by the new *Nelson-Denny Reading Test* CD- ROM. This study addressed three aspects of validity, score comparability between reading rate measured by the computer version and the paper and pencil version, concurrent validity with reading comprehension and predictive validity with academic achievement. Concurrent validity was studied through a score equivalence special study with 182 students taking both the computerized and paper and pencil versions. Mean score differences were not statistically significant and correlations were significant at the .05 level. Construct validity was displayed through a curvilinear relationship with reading comprehension in addition to adding a significant increase to predicted variance for reading comprehension after reading vocabulary was first used as a predictor. Predictive validity was demonstrated by significant additions to the percent of variance explained in the prediction of high school and English course grades after the reading total score was first used as a predictor. The results of this study support the use of this computerized measure of reading rate and suggest several areas of possible usefulness in the understanding of reading comprehension and the prediction of academic achievement.

The validity of a computerized measure of reading rate

Recent improvements in computer technology and psychometrics have encouraged the delivery of educational and psychological tests through computers. The advantages which are frequently cited for computerized testing include immediate feedback of results, increased examinee interest, and reduced costs of production, administration, and scoring. Many large and small scale testing programs have transitioned to computer delivery of their tests. However, when tests are converted to computer delivery, the validity of the measures cannot be assumed to be the same as for their paper and pencil counterparts. As a result, the validity of the computerized measures needs to be separately established. The present study addressed the validity of a computerized measure of reading rate. This investigation focused on three areas; concurrent validity with a paper and pencil measure of reading rate, construct validity with reading comprehension and predictive validity for student grades.

Concurrent validity. One aspect of establishing the validity of any measure is to show that the measure is related to other measures of the same construct. In this study, the focus was on how well the new computerized reading rate correlated with the reading rate measure in the paper and pencil versions of the same test. In two reviews of computer-paper and pencil equivalence studies, Mazzeo and Harvey (1988) and Mead and Drasgow (1993) pointed out the that any differences in stimuli delivery systems and/or response procedures in the two formats could threaten their score equivalence. In fact, concern about computer delivery of reading texts initially focused on the replacement of paper with computer screens as the delivery medium. In the research during the 1980's the focus of this research was the legibility of the computer screen

defined by Reinking (1988) as "how physical characteristics of written text affect factors such as visual fatigue, reading speed, and comprehension." Several studies found that texts displayed electronically were read more slowly than conventional texts (Gould & Grischkowsky, 1983) However, other research suggested that this slower rate was related to variables that have improved in recent years such as the size of character fonts and screen resolution (Gould, Alfaro, Finn, Haupt & Minuto, 1987). More recent studies have not shown large differences in reading rate between text on computer screens and paper (Clansing & Schmitt, 1990).

In the present study, there was a differences in the manner in which reading rate was measured in the two formats. In the paper and pencil format, students marked the **line** they were reading after one minute of reading, while in the computerized version the students clicked on the **word** they were reading. As a result, the word count used as the reading rate measure in the pencil and paper version was the *middle* word in the last line read while in the computerized version the measure was the last word read.

Construct validity: relationship with comprehension. One step in demonstrating construct validity is to use accepted theory to generate and test hypotheses about how different constructs are related. The relationship between speed of processing and cognitive ability has long been a research subject in psychology (Furneaux, 1960; Carroll, 1993). In reading research, this is research subject is operationalized by the relationship between reading rate and reading comprehension (Carver, 1990). Reading comprehension is the result of an interaction between incoming read text and the reader's network of stored knowledge in long-term memory. Working memory has been called the "cognitive workbench" where the incoming information is integrated with the stored knowledge.

These models usually assume limited working memory capacity for performing storage and processing tasks. As a result, recognizing words very slowly would take up more of the total capacity and create difficulties in other processes and recognizing words very quickly would free up more capacity for integrating the text and the stored knowledge. Therefore, differences in reading rate could affect reading comprehension.

However, researchers have not found a simple relationship between reading comprehension and reading rate. Carver (1982) forced students to read text at different rates. He found that reading comprehension decreased at very slow and very fast reading rates. Although this research suggested a curvilinear relationship between reading rate and comprehension, Carver did not attempt to model the relationship mathematically. Rankin (1993) divided students into high comprehension-high speed, high comprehension-low speed, low comprehension-low speed, and low comprehension and --high speed. While she found clear differences between high and low comprehension students on a series of information processing tasks, the pattern of results was less clear for high and low speed readers.

Predictive validity. A primary use of Nelson-Denny test scores is to inform student placement decisions into appropriate reading and/or language arts classes. As a result, if reading rate contributes to the prediction of course grades even after considering the vocabulary and reading comprehension scores, then this would be important evidence of predictive validity. Although very little research has been published on this topic, Lord (1956) and more recently Wilson (1989) both concluded that speededness was an important predictor of course grades. Lord (1956) found that highly speeded vocabulary tests were the best single predictor of course grades for 600 Naval Academy students.

Wilson (1989) created a speed of comprehension score separate from level of comprehension from the verbal section of the Graduate Record Examination. He found that the speed score had higher correlations with course grades for most students than the level of comprehension score.

The present study examined the concurrent, construct and predictive validity of reading rate as measured by the computerized *Nelson Denny Reading Test*. Concurrent validity was studied by comparing paper and pencil and computerized reading rate measures for the same students. Construct validity was investigated through linear and curvilinear mathematical models to examine the relationship of reading rate and reading comprehension. Predictive validity was studied through the contribution of reading rate to prediction of course grades.

Method

Subjects. In the fall and winter of 1999-2000, students participated in this study as part of a project concerning the introduction of the computerized version of the *Nelson Denny Reading Test*. For the primary part of this study, high schools and two-year college were randomly assigned to administer Form G or Form H to their students. As a result of this design, random assignment was at the school level for this study. However, because of concerns early in the study about participation, students from four-year colleges were disproportionately assigned Form G. Over thirty high schools, two-year and four-year colleges administered the computerized versions of the *Nelson Denny* in return for a stipend and score reports. As a result, over 3,000 students were administered the computerized versions of the *Nelson Denny*.

To study the equivalence of paper and pencil and computerized reading rate measures, a subgroup of students were administered both the paper and pencil and computerized versions of the *Nelson Denny*. A total of six institutions volunteered to participate and were randomly assigned to administer one of two computer forms and the alternative paper and pencil form. Because computers are limited in number at most educational institutions, the data collection for the score equivalence part of this study needed a design that minimized the number of students required.

The random groups counterbalanced design requires fewer examinees for the same level of accuracy as other data collection designs. In the counterbalanced design the same students take both forms of the test but are assigned to take the forms in different orders to balance any practice effect from the order of the tests. Students were requested to take both forms, G and H, in different formats within a period of two weeks. Schools were randomly assigned to either the computerized form G - paper and pencil form H group or the computerized form H - paper and pencil form G group. Schools were given instructions to randomly assign their students to two testing orders; either administering the computerized form first or the paper and pencil form first. The participating schools in the special study included one high school, two community colleges and two four- year colleges. These schools provided both computer based and paper and pencil test scores for over 200 students.

Instrument. The *Nelson Denny Reading Test* is a widely used reading comprehension test with high reliability and established validity (Brown, Fishco, & Hanna, 1993). The test provides three reading achievement scores including vocabulary, reading comprehension and a total score as well as a reading rate score. In the most

recent forms, G and H, there were 80 vocabulary items each with five answer choices and a time limit of 15 minutes. The first passage of the reading comprehension section was used to determine reading rate by prompting the examinee to click or key on the last word read when prompted after one minute of time.

After students were administered the computerized form of the *Nelson Denny*, they answered 19 survey questions about themselves and the experience of computerized testing. Two of these questions asked them about their high school grade-point-average and the grade in their last English/Language Arts course. These questions were "Which category best describes your high school grade point average? A (3.4 to 4.), B (2.7 to 3.3), C (1.7 to 2.6), D (0.7 to 1.6), F (0.0 to 0.6)" and "What grade did you receive in the last English or Literature class you took? A, B, C, D, F". Past research has supported the validity of grades as reported by students on surveys (Freeberg, Rock, & Pollack, 1989)

Participating schools needed to have Windows 95, 98 or NT/2000 IBM compatible personal computers with 32 MG of RAM with 256 colors in a low resolution mode. The *Nelson Denny* CD ROM program automatically set the computer screen to a 800 by 600 resolution. After an informal pilot test, the text and questions were set to a size of 12 point in a Times New Roman font. Lastly, a mouse was required with the personal computers because the computerized program had no keyboard functionality.

Analyses: *Score equivalence*. For the study of the comparability of the reading rates of the paper and pencil and the computer versions, the analyses were performed on the student scores from the counterbalanced design. With a counterbalanced design, a concern is that taking one form of a test closely followed by another form would effect the scores for the second form. If the effect produced higher scores on the second form it

would be considered a practice effect and if the scores on the second form were consistently lower, a fatigue effect. As a result, scores were first examined for the four order by form groups separately. Practice effects were defined as consistent score advantages or disadvantages statistically significant at the .05 level for the form administered last. To test the score equivalence of these pairs of scores, t-test's for dependent scores were performed for the scores in each group.

Construct validity. For the study of the relationship of reading rate and comprehension, scatter plots of reading rate and reading comprehension were used to guide the fitting of linear and quadratic components to model the relationship. Stepwise regression analysis with forced entry was then used to test the significance of the information added to the prediction of reading comprehension after the influence of vocabulary was accounted for. The statistical significance of each increment in the proportion of variance explained was tested with an F ratio.

Predictive validity. The last part of the study addressed the predictive validity of the reading rate measure. The criterion variables in these linear regression analyses were self-reported English course grades and grade-point-averages. The significance of the reading rate variable was tested after the total reading (comprehension plus vocabulary) score was entered first into the model used to predict student grades. For both analyses, the increase in the R-squared or the proportion of variance accounted for, was tested for statistical significance at the .05 level.

Results

Table 1 displays the descriptive statistics for the reading rate scores from students who took both the computerized and paper and pencil versions of the *Nelson Denny*.

This table displays the number of students who took each set of forms, the mean and standard deviation of the scores on each format, the correlation of those scores and the statistical significance of the score differences. The scores are reported in scale scores to adjust for the small difficulty differences between the Forms G and H. As mentioned before, incentives were awarded to schools and not students. As a result, before the analyses were conducted the veracity of the individual scores needed to be established. After deletion of aberrant scores identified through outlier statistics and inspection of score distributions, 182 students in the special study had scores for both the computer and paper and pencil forms.

The computerized and paper and pencil score differences were very small for the computer version of Form G (.92) and slightly larger for the computer version of Form H (4.39). However, the last column in Table 1 shows that neither difference was statistically significant at the .05 level. Both correlations were statistically significant at the .001 level but somewhat low for parallel measures. The magnitude of the correlations was probably deflated by the different methods used to measure reading rate on the computerized and paper and pencil versions of the test.

Table 2 displays the means and standard deviations for the scores used in the rest of this study. The last English grade and high school grade point average were scaled "5=A, 4=B, 3=C, 2=D and 1=F". Scores for vocabulary, comprehension, total (vocabulary + comprehension) and reading rate are reported in the *Nelson Denny* scale scores which have a mean of 200 and a standard deviation in the 1993 population used for norming the paper and pencil forms. After screening the data using outlier statistics and inspection of score distributions, 3,282 students high school and college students who

took Form G or Form H of the computerized forms of the *Nelson Denny* were included in the rest of the study. These scores indicate that as expected most of the scores were close to a mean of 200 and a standard deviation of 25. The two self-reported grade variables indicate that on average these students reported 'B' high school grade point averages and grades in their last English course.

Figures 1 and 2 display scatterplots for reading rate and reading comprehension scores. Figure 1 shows the results of a simple linear relationship (the straight line) between the two variables and Figure 2 displays a nonlinear quadratic relationship (the curvilinear line). The R-squared shown to the left of the figure for these results is .01 (r = .11). In contrast, Figure 2 shows that the R-squared result for the quadratic relationship was .15 (r = .39). Because the results in Figures 1 and 2 suggest a curvilinear relationship, a squared reading rate variable was included in the prediction models. To decrease collinearity, reading rate was first standardized into z-scores and then squared for entry into the model after reading rate (Kleinbaum, Kupper and Muller, 1988).

Table 3 shows the results for the stepwise regression model in which vocabulary and reading rate were used to predict reading comprehension. This table shows the standardized regression coefficients, the single and multiple correlations, the adjusted R-squared or percent of variance explained, the increase in R-squared for each variable, and the statistical significance of the additional prediction for each variable. The focus of this model was the prediction added by reading rate after vocabulary was first used to predict reading comprehension.

The multiple correlation (R) in Table 3 showed that vocabulary was significantly related to reading comprehension. The single correlation was .716 with an adjusted R-

squared of .513. Vocabulary predicted about 50% of the variance of reading comprehension scores. The model showed a nonsignificant increase in the multiple correlation when reading rate was added to the model. However, when the quadratic reading rate term was added to the prediction model a statistically significant increase in the multiple correlation was displayed.

Table 4 shows the results for the stepwise regression model in which the total reading score and reading rate were used to predict self reported high school grade point average. The table shows the standard regression coefficients, the single and multiple correlations, the adjusted R-squared or percent of variance explained, the increase in R-squared for each variable, and the statistical significance of the additional prediction for each variable. The focus of this model was the prediction added by reading rate after total score was first used to predict high school grade point average.

The multiple correlation (R) in Table 4 showed that total score was significantly related to high school grade point average. The single correlation was .307 with an adjusted R-squared of .094. Total score predicted about 9% of the high grade point variance. The model showed a significant increase in the multiple correlation when reading rate was added to the model and a nonsignificant increase in the R-squared when the quadratic reading rate term was added to the prediction models.

Table 5 shows the results for the stepwise regression model in which the total reading score (vocabulary + comprehension) and reading rate were used to predict the grade in the last English course taken. The focus of this model was the prediction added by reading rate after total score was first used to predict last grade in an English course. The multiple correlation (R) in Table 5 showed that total score was significantly related

to last grade in an English course. The single correlation was .301. Total score predicted about 9% of the last grade in an English course variance. The model showed significant increases in the multiple correlation when reading rate and reading rate squared were added to the model.

Summary

The present study investigated the validity of a computerized measure of reading rate. The results showed concurrent validity through nonsignificant differences between the computerized measure of reading rate and a paper and pencil reading rate measure. Construct validity was established through a curvilinear relationship with reading comprehension which is consistent with reading theory. In addition, even after vocabulary is considered, the reading rate squared variable added a significant increase to the prediction of reading comprehension. Predictive validity was shown by statistically significant increases to percent of variance explained for high school grade point average and grade received in last English course. Statistically significant increases in predicted variance were shown for both grade variables by the reading rate variable and for the last English course grade for only the reading rate variable. Further studies are needed to determine the practical significance of these increases for placement decisions.

This study has educational importance because it studied in depth the validity of a reading rate measure for a popular and useful test of reading skills. The computer version is easier to administer and provides results faster than paper and pencil version. The results provide validity evidence for teachers and administrators who use the test for diagnostic and placement purposes. In addition, the study contributes to the understanding of the relationship of reading rate and comprehension, the comparability of

computer and paper and pencil measures of reading rate, and the usefulness of reading rate in predicting academic achievement.

References

Brown, J. I., Fishco, V. V., & Hanna, G. (1993). Nelson-Denny Reading Test: Technical Report Forms G&H. Chicago: The Riverside Publishing Company.

Carroll, J. B. (1993). Human cognitive abilities: A survey of factor-analytic studies. New York: Cambridge University Press.

Carver, R. P. (1982). Optimal rate of reading prose. Reading Research Quarterly, 18, 56-88.

Clansing, C., & Schmitt, D. (1990). Paper versus CRT: Are reading rate and comprehension affected? Paper available from ERIC.

Freeberg, N. E., Rock, D. A., & Pollack, J. (1989). Analysis of the Revised Student Descriptive Questionnaire: Phase II, Predictive validity of academic self report. (College Board Report No 89-8). New York: College Entrance Examination Board.

Furneaux, W. D. (1960). Intellectual abilities and problem solving behaviour. In Eysenck, H. J. (ed.) Handbook of Abnormal Psychology. London: Pittman.

Gould, J. D., & Grischkowsky, N. (1983). Doing the same work with paper and cathode ray tube displays (CRT). Human Factors, 24, 329-338.

Gould, J. D., Alfaro, L., Finn, R., Haupt, B., & Minuto, A. (1987). Reading from CRT displays can be as fast as reading from paper. Human Factors, 29(5), 497-517.

Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). Applied regression analysis and other multivariate methods. Boston: PWS-KENT Publishing Company.

Lord, F. M. (1956). A study of speed factors in tests and grades. Psychometrika, 21, 31-49.

Mazzeo, J., & Harvey. A. L. (1988). The equivalence of scores from automated and conventional educational and psychological tests: A review of the literature. (College Board Report No 91-5). Princeton NJ: Educational Testing Service.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper and pencil cognitive ability tests: A meta-analysis. Psychological Bulletin, 114, 449-458.

Rankin, J. (1993). Information processing differences of college age readers differing in reading comprehension and speed. Journal of Reading Behavior, 25 (3), 261-278.

Reinking, D. (1988). Computer mediated text and comprehension differences: The role of reading time, reader preference, and estimation of learning. Reading Research Quarterly, 13 (4), 486-88.

Wilson, K. M. (1989). Population differences in speed versus level of GRE reading comprehension: An exploratory study. ETS Research Report 89-36. Princeton, NJ: ETS.

Table 1. Descriptive scaled score statistics for reading rate by form for special study

		Computer		Paper			
Order	N	Mean	SD	Mean	SD	Score Correlation	Difference Significance
CD G – PP H	84	192.74	26.70	192.26	22.76	.54	.856
CD H – PP G	101	204.42	34.72	208.71	39.41	.66	.166

Table 2. Descriptive score statistics from norms study

	N= 3,282	
Variable	Mean	SD
Vocabulary	198.63	23.14
Comprehension	201.09	23.98
Total Score	199.88	23.03
Last English Grade [a]	3.97	.80
High School GPA [a]	3.97	.90
Reading Rate	204.85	33.12

[a] Grade Scale was "5=A, 4=B, 3=C, 2=D, 1=F".

Table 3. Contribution of reading rate to prediction of reading comprehension

Prediction Model	Standardized coefficient	R	R squared	Increase in R squared	Statistical significance
Vocabulary	.716	.716	.513	.513	<.001
Vocabulary, reading rate	.715 .006	.716	.513	.000	.616
Vocabulary, reading rate, reading rate^2	.635 .163 -.228	.735	.540	.028	<.001

Table 4. Contribution of reading rate to prediction of high school grade point average

Prediction Model	Standardized coefficient	R	Adjusted R squared	Increase in R squared	Statistical significance
Total score	.307	.307	.094	.094	<.001
Total score, reading rate	.300 .054	.312	.097	.003	<.001
Total score, reading rate, reading rate^2	.288 .086 -.044	.313	.097	.001	.073

Table 5. Contribution of reading rate to prediction of grade for last English or Literature class

Prediction Model	Standardized coefficient	R	Adjusted R squared	Increase in R squared	Statistical significance
Total score	.301	.301	.090	.091	<.001
Total score, reading rate	.294 .050	.305	.093	.002	.003
Total score, reading rate, reading rate^2	.277 .099 -.067	.308	.094	.002	.006

Figure 1. Reading rate and comprehension - Linear model

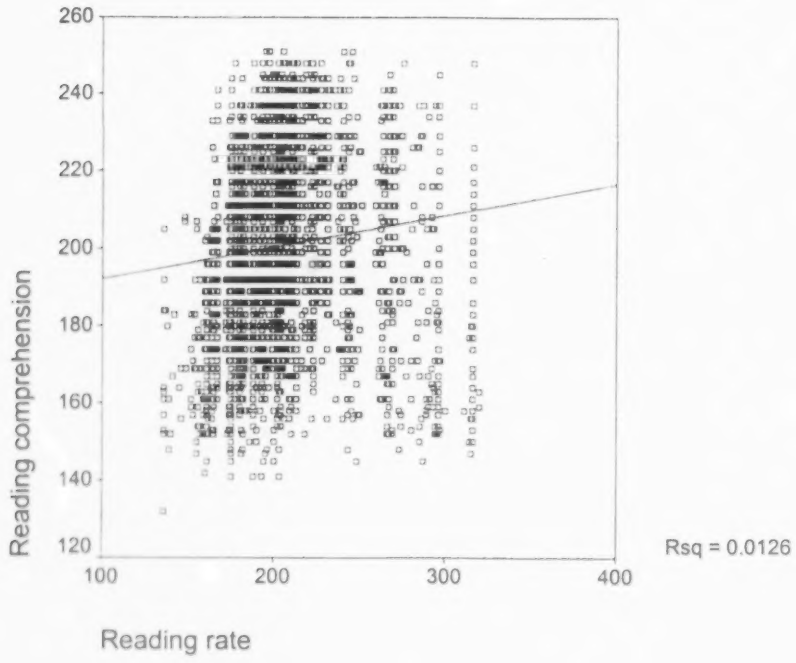
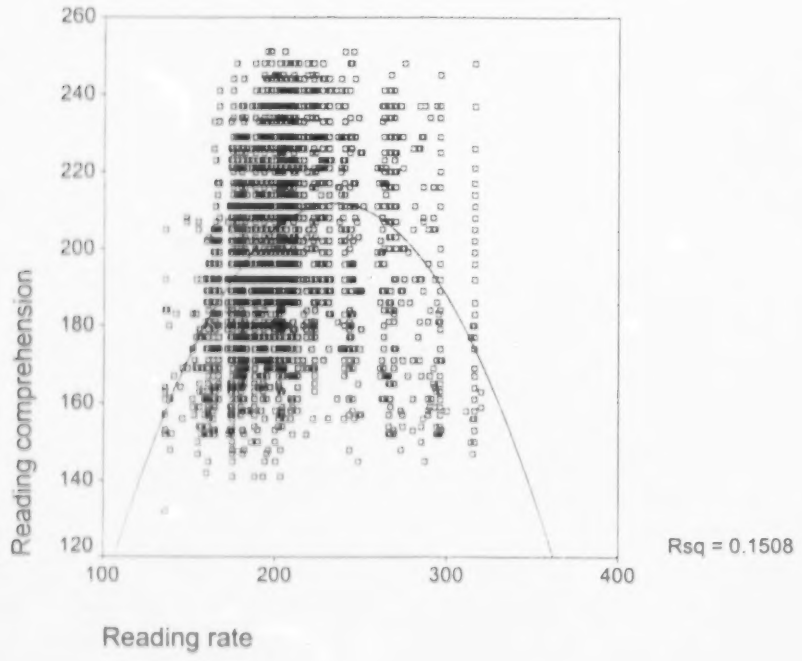


Figure 2. Reading rate and comprehension - Curvilinear model



BEST COPY AVAILABLE